

3 Multipla korespondenčna analiza

Enostavna oz. bivariatna korespondenčna analiza, ki smo jo podrobno prikazali v 2. poglavju, je namenjena analizi odnosov med dvema nominalnima spremenljivkama. Multipla korespondenčna analiza pa predstavlja naravno posplošitev bivariatne analize na primer treh ali več nominalnih spremenljivk.

3.1 Burtova tabela

Predpostavimo, da želimo proučiti odnose med Q nominalnimi spremenljivkami z multiplo korespondenčno analizo. Izhodišče analize je Burtova tabela \mathbf{B} (Burt C., 1950), to je kvadratna simetrična matrika reda J (3.2) z naslednjo bločno obliko:

$$\mathbf{B} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \cdots & \mathbf{F}_{1Q} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \cdots & \mathbf{F}_{2Q} \\ \vdots & \vdots & & \vdots \\ \mathbf{F}_{Q1} & \mathbf{F}_{Q2} & \cdots & \mathbf{F}_{QQ} \end{bmatrix} \quad (3.1)$$

sestavljen iz Q^2 podmatrik z naslednjim pomenom:

- podmatrike $\mathbf{F}_{qq} = \mathbf{D}_{f_q}$ ($q = 1, 2, \dots, Q$) na glavni diagonalni matrike \mathbf{B} so diagonalne matrike reda J_q , katerih diagonalni elementi so absolutne frekvence $f_{\cdot j_q}$ ($j_q = 1, 2, \dots, J_q$), ki se nanašajo na vrednosti spremenljivk V_q ($q = 1, 2, \dots, Q$).
- podmatrike $\mathbf{F}_{qq''}$ ($q \neq q''$; $q, q'' = 1, 2, \dots, Q$), ki ne pripadajo glavni diagonalni matrike \mathbf{B} , so dvorazsežne kontingenčne tabele in izražajo povezanost med vrednostmi nominalnih spremenljivk V_q in $V_{q''}$ ($q \neq q''$; $q, q'' = 1, 2, \dots, Q$).

Skupno število vrednosti proučevanih nominalnih spremenljivk J je enako

$$J = \sum_{q=1}^Q J_q \quad (3.2)$$

Vsota stolpcev matrike \mathbf{B} (ali vrstic, kajti \mathbf{B} je simetrična matrika) je enaka vektorju

$$\mathbf{B}\mathbf{1} = Q\mathbf{f} \quad (3.3)$$

pri čemer je vektor \mathbf{f} enak

$$\mathbf{f} = \begin{bmatrix} \overline{\mathbf{f}_1} \\ \overline{\mathbf{f}_2} \\ \vdots \\ \overline{\mathbf{f}_Q} \end{bmatrix} \quad (3.4)$$

njegovi elementi pa so absolutne frekvence $f_{.j_q}$ ($j_q = 1, 2, \dots, J_q$; $q = 1, 2, \dots, Q$), ki se nanašajo na vrednosti spremenljivk V_q ($q = 1, 2, \dots, Q$).

Vsota vseh elementov matrike \mathbf{B} je enaka

$$\mathbf{1}'\mathbf{B}\mathbf{1} = \mathbf{1}'Q\mathbf{f} = Q\mathbf{1}'\mathbf{f} = Q^2H. \quad (3.5)$$

Burtovi tabeli \mathbf{B} običajno pridružimo zbirno vrstico in zbirni stolpec $Q\mathbf{f}$ (3.3) ter vsoto vseh njenih elementov Q^2H (3.5), da dobimo razširjeno Burtovo tabelo, ki ima naslednjo bločno obliko:

$$\left[\begin{array}{c|c} \mathbf{B} & Q\mathbf{f} \\ \hline Q\mathbf{f}' & Q^2H \end{array} \right] \quad (3.6)$$

3.2 Matrika relativnih frekvenc Burtove table

Če Burtovo tabelo \mathbf{B} (3.1) delimo z vsoto vseh njenih elementov Q^2H (3.5), dobimo matriko relativnih frekvenc:

$$\mathbf{P}_B = \frac{1}{Q^2H} \mathbf{B} \quad (3.7)$$

Podobno, če razširjeno Burtovo tabelo (3.6) delimo z vsoto vseh njenih elementov Q^2H (3.5), dobimo razširjeno matriko relativnih frekvenc:

$$\left[\begin{array}{c|c} \mathbf{P}_B & \mathbf{p}_B \\ \hline \mathbf{p}_B' & 1 \end{array} \right] \quad (3.8)$$

pri čemer je

$$\mathbf{p}_B = \frac{Q\mathbf{f}}{Q^2H} = \frac{1}{QH}\mathbf{f} \quad (3.9)$$

vektor marginalnih relativnih frekvenc oz. povprečni profil.

3.3 Matrika profilov Burtove tabele

Pomen posameznih skupin je v Burtovi tabeli \mathbf{B} predstavljen z absolutnimi frekvencami. Da bi dobili boljši vpogled v odnose med skupinami, običajno izračunamo pogojne relativne frekvence po vrsticah (ali po stolpcih!) Burtove tabele. Če torej delimo po vrsticah elemente Burtove tabele (3.1) z elementi njenega zbirnega stolpca $Q\mathbf{f}$ (3.3) dobimo matriko pogojnih relativnih frekvenc oz. matriko vrstičnih profilov

$$[\text{diag}(Q\mathbf{f})]^{-1}\mathbf{B} = \frac{1}{Q}\mathbf{D}_f^{-1}\mathbf{B} \quad (3.10)$$

pri čemer je \mathbf{D}_f naslednja bločna diagonalna matrika

$$\mathbf{D}_f = \begin{bmatrix} \mathbf{D}_{f_1} & & & \mathbf{0} \\ & \mathbf{D}_{f_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{D}_{f_q} \end{bmatrix} \quad (3.11)$$

Če pa delimo elemente zbirne vrstice Burtove tabele $Q\mathbf{f}$ (3.3) z vsoto vseh elementov Burtove tabele Q^2H (3.5) dobimo profil zbirne vrstice oz. že omenjeni povprečni profil \mathbf{p}_B (3.9).

Na podlagi matrike vrstičnih profilov (3.10) lahko oblikujemo razširjeno matriko vrstičnih profilov Burtove tabele

$$\left[\begin{array}{c|c} \frac{1}{Q}\mathbf{D}_f^{-1}\mathbf{B} & \mathbf{1} \\ \hline \mathbf{p}'_B & 1 \end{array} \right] \quad (3.12)$$

Zaradi simetričnosti Burtove tabele so matrike stolpičnih profilov enake transponiranim matrikam vrstičnih profilov.

3.4 Določanje optimalnega podprostora

Vzemimo, podobno kot v primeru enostavne korespondenčne analize, kot izhodišče za izračun koordinat točk in pripadajočih algebrskih kazalcev korespondenčne analize matriko

$$\mathbf{D}_B^{-1/2}(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}_B') \mathbf{D}_B^{-1/2} \quad (3.13)$$

Matrika $(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}_B')$ je matrika hkrati centriranih vrstic in stolpcev matrike relativnih frekvenc Burtove tabele \mathbf{P}_B (3.7) z rangom

$$\text{rang}(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}_B') \leq L \quad (3.14)$$

pri čemer je

$$L = J - Q \quad (3.15)$$

in je J skupno število vrednosti vseh Q spremenljivk (3.2), L pa število netrivialnih glavnih komponent.

Matrika

$$\mathbf{D}_B = \frac{Q}{Q^2 H} \mathbf{D}_f = \frac{1}{QH} \mathbf{D}_f \quad (3.16)$$

je diagonalna matrika uteži vrstic (oz. stolpcev) Burtove matrike, njeni diagonalni elementi so pravzaprav elementi vektorja \mathbf{p}_B (3.9).

S spektralno dekompozicijo matrike (3.13) izračunamo diagonalno matriko glavnih inercij \mathbf{D}_λ in matriko standardiziranih koordinat glavnih komponent $\mathbf{\Gamma}$, pri čemer je $\mathbf{\Gamma}' \mathbf{D}_B \mathbf{\Gamma} = \mathbf{I}$.

Izračun izvedemo v naslednjih dveh korakih:

1. Najprej izračunamo spektralno dekompozicijo matrike $\mathbf{D}_B^{-1/2}(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}_B') \mathbf{D}_B^{-1/2}$ (3.13)

$$\mathbf{D}_B^{-1/2}(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}_B') \mathbf{D}_B^{-1/2} = \tilde{\mathbf{\Gamma}} \mathbf{D}_\lambda \tilde{\mathbf{\Gamma}}' \quad (3.17)$$

pri čemer je

$$\tilde{\mathbf{\Gamma}}' \tilde{\mathbf{\Gamma}} = \mathbf{I} \quad (3.18)$$

in \mathbf{D}_λ diagonalna matrika glavnih inercij λ_j

$$\mathbf{D}_\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_L) \quad (3.19)$$

2. Matriko standardiziranih koordinat glavnih komponent $\mathbf{\Gamma}$ izračunamo

takole:

$$\mathbf{\Gamma} = \mathbf{D}_B^{-1/2} \tilde{\mathbf{\Gamma}} \quad (3.20)$$

Na podlagi spektralne dekompozicije matrike (3.13) ugotavljamo, da je aproksimacija prvotnega prostora z optimalnim podprostorom odvisna od vseh vrednosti Burtove tabele, torej tudi od diagonalnih podmatrik \mathbf{F}_{qq} ($q=1,2,\dots,Q$). Podmatrike \mathbf{F}_{qq} ne izražajo povezanosti med po dvema različnima spremenljivkama, ob tem pa močno povečajo vrednost skupne inercije in tako zmanjšajo deleže glavnih inercij λ_j v skupni inerciji In . Zato J.P. Benzécri (1979) predlaga, da se kot analitično pomembne obravnava le tiste glavne komponente, katerih inercije so višje od $1/Q$. Nove, bolj realistične glavne inercije, imenovali jih bomo modificirane inercije, izračunamo takole

$$\tilde{\lambda}_j = \left[\frac{Q}{Q-1} \left(\lambda_j - \frac{1}{Q} \right) \right]^2, \quad j=1,2,\dots,X \quad (3.21)$$

Na podlagi X modificiranih inercij lahko definiramo množico nestandardiziranih glavnih komponent, izraženih z matriko \mathbf{Y} , osnova zanjo pa je prvih X standardiziranih glavnih komponent matrike $\mathbf{\Gamma}$

$$\mathbf{Y} = \mathbf{\Gamma} \mathbf{D}_{\tilde{\lambda}}^{1/2} \quad (3.22)$$

pri čemer je $\mathbf{D}_{\tilde{\lambda}}$ diagonalna matrika modificiranih inercij $\tilde{\lambda}_j$

$$\mathbf{D}_{\tilde{\lambda}} = \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_X) \quad (3.23)$$

Burtova tabela \mathbf{B} je po svoji strukturi analogna variančno-kovariančni matriki s Q intervalnimi spremenljivkami. V tem smislu vsaki podmatriki $\mathbf{F}_{qq''}$ ($q, q''=1,2,\dots,Q$) (dvorazsežni kontingenčni tabeli) matrike \mathbf{B} odgovarja kovarianca dveh intervalnih spremenljivk. Klasična multivariatna analiza intervalnih spremenljivk običajno predpostavlja multivariatno normalno porazdelitev spremenljivk. Njeno izhodišče so zato običajno momenti drugega reda. Analogno, korespondenčna analiza z izhodiščem v Burtovi tabeli prav tako temelji na povezanostih, ki nastopajo med vsemi možnimi dvojicami nominalnih spremenljivk. Zanj je namreč značilno, da hkrati obravnava odnose med vsemi dvorazsežnimi kontingenčnimi tabelami. Zato je multipla

korespondenčna analiza pravzaprav bolj "večkratno-bivariatna" kot pa multivariatna analitična metoda (de Leeuw, 1973).

Primer 2

Naše izhodišče bodo podatki zbrani v okviru socialno-ekonomske raziskave "Pogoji življenja in težnje Francozov" (L. Lebart in Y. Houzel van Effentere, 1980). Podatki se nanašajo na reprezentativni delni vzorec 1000 anketirancev in na izbranih sedem spremenljivk ($Q = 7$) s skupno 25 vrednostmi ($J = 25$):

- spol (SPOM - moški ; SPOŽ - ženski),
- najvišja stopnja dosežene šolske izobrazbe (IZO1 - brez dokončane osnovne šole , IZO2 - dokončana osnovna šola ; IZO3 - nedokončana srednja šola ; IZO4 – dokončana srednja šola ; IZO5 - vsaj nedokončana fakulteta)
- stanovanjske razmere (STN1 - hipotečni dolžnik ; STN2 - lastnik ; STN3 - najemnik ; STN4 - brezplačno bivanje)
- lastništvo delnic ali obveznic (DEL1 - da ; DEL2 - ne)
- lastništvo nepremičnin, brez stanovanja, v katerem biva (NEP1 - da ; NEP2 - ne)
- starost (STR1 - 19 do 24 let ; STR2 - 25 do 34 let ; STR3 - 35 do 49 let ; STR4 – 50 do 64 let ; STR5 - 65 let in več)
- velikost kraja bivanja (število prebivalcev) (VEL1 - pod 2.000 ; VEL2 – 2000 do pod 20000 ; VEL3 - 20000 do pod 100000 ; VEL4 – 100000 do pod 500000 ; VEL5 - nad 500000).

Na podlagi vrednosti proučevanih spremenljivk lahko za vzorec anketiranih oseb oblikujemo razširjeno Burtovo tabelo (3.6), prikazano v tabeli 3.1 .

Pomen posameznih skupin je v Burtovi tabeli predstavljen z absolutnimi frekvencami. Da bi dobili boljši vpogled v odnose med skupinami izračunajmo pogojne relativne frekvence po vrsticah Burtove tabele (3.10) oz. matriko vrstičnih profilov, prikazano v tabeli 3.2 (zaradi lažje berljivosti smo pogojne relativne frekvence izrazili v odstotkih). Tabela vključuje tudi zbirno vrstico \mathbf{p}_B (3.9), ki prestavlja povprečni profil.

Tabela 3.1: Razširjena Burtova tabela (Vir: L. Lebart et al., 1984, stran 104).

	SPO					IZO					STN					DEL					NEP					STR					VEL					Skupaj
	M	Z	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5				
SPOM	469	0	102	164	69	65	69	62	151	224	32	54	415	45	434	18	111	169	84	87	42	40	81	161	145	3283										
SPOZ	0	531	94	163	91	102	81	58	142	302	29	67	464	47	484	22	136	187	104	82	41	47	94	168	181	3717										
IZO1	102	94	196	0	0	0	0	17	59	105	15	12	184	11	185	9	27	54	47	59	19	17	35	70	55	1372										
IZO2	164	163	0	327	0	0	0	45	116	151	15	26	301	21	306	0	55	125	84	63	43	34	67	110	73	2289										
IZO3	69	91	0	0	160	0	0	16	37	98	9	19	141	14	146	14	47	60	27	12	12	13	34	52	49	1120										
IZO4	65	102	0	0	0	167	0	27	39	87	14	27	140	19	148	16	62	63	16	10	8	15	21	57	66	1169										
IZO5	69	81	0	0	0	0	150	15	42	85	8	37	113	17	133	1	56	54	14	25	1	8	18	40	83	1050										
STN1	62	58	17	45	16	27	15	120	0	0	0	11	109	7	113	3	23	68	20	6	7	20	27	48	18	840										
STN2	151	142	59	116	37	39	42	0	293	0	0	60	233	48	245	9	24	91	80	89	59	33	62	72	67	2051										
STN3	224	302	105	151	98	87	85	0	0	526	0	45	481	23	503	22	180	182	79	63	11	29	80	191	215	3682										
STN4	32	29	15	15	9	14	8	0	0	0	61	5	56	4	57	6	20	15	9	11	6	5	6	18	26	427										
DEL1	54	67	12	26	19	27	37	11	60	45	5	121	0	39	82	2	18	35	27	39	4	9	22	36	50	847										
DEL2	415	464	184	301	141	140	113	109	233	481	56	0	879	43	836	38	229	321	161	130	79	78	153	293	276	6153										
NEP1	35	47	11	21	14	19	17	7	48	23	4	39	43	82	0	3	11	27	20	21	7	12	12	25	26	574										
NEP2	434	484	185	306	146	148	133	113	245	503	57	82	836	0	918	37	236	329	168	148	76	75	163	304	300	6426										
STR1	18	22	9	0	14	16	1	3	9	22	6	2	38	3	37	40	0	0	0	0	2	4	8	15	11	280										
STR2	111	136	27	55	47	62	56	23	24	180	20	18	229	11	236	0	247	0	0	0	13	16	34	102	82	1729										
STR3	169	187	54	125	60	63	54	68	91	182	15	35	321	27	329	0	0	356	0	0	31	34	71	100	120	2492										
STR4	84	104	47	84	27	16	14	20	80	79	9	27	161	20	168	0	0	0	188	0	25	20	26	60	57	1316										
STR5	87	82	59	63	12	10	25	6	89	63	11	39	130	21	148	0	0	0	0	169	12	13	36	52	56	1183										
VEL1	42	41	19	43	12	8	1	7	59	11	6	4	79	7	76	2	13	31	25	12	83	0	0	0	0	581										
VEL2	40	47	17	34	13	15	8	20	33	29	5	9	78	12	75	4	16	34	20	13	0	87	0	0	0	609										
VEL3	81	94	35	67	34	21	18	27	62	80	6	22	153	12	163	8	34	71	26	36	0	0	175	0	0	1225										
VEL4	161	168	70	110	52	57	40	48	72	191	18	36	293	25	304	15	102	100	60	52	0	0	0	329	0	2303										
VEL5	145	181	55	73	49	66	83	18	67	215	26	50	276	26	300	11	82	120	57	56	0	0	0	0	0	2282										
Skupaj	3283	3717	1372	2289	1120	1169	1050	840	2051	3682	427	847	6153	574	6426	280	1729	2492	1316	1183	581	609	1225	2303	2282	49000										

Burtova tabela **B** je simetrična bločna matrika, ki združuje vse možne pare dvorazsežnih kontingenčnih tabel $F_{qq'}$ ($q, q' = 1, 2, \dots, Q$) neke množice nominalnih spremenljivk. Če usmerimo našo pozornost k njenim posameznim kontingenčnim tabelam, lahko za vsako izmed njih izračunamo statistiko χ^2 ⁸⁾, ki meri povezanost med dvema nominalnima spremenljivkama, in ustrezno vrednost P (točno stopnjo značilnosti) ter jih prikažemo v tabeli 3.3.

Tabela 3.3 Prikaz vrednosti statistike χ^2 in vrednosti P za vse možne pare nominalnih spremenljivk (Vir: Tabela 3.1).

χ^2 P	SPO	IZO	STN	DEL	NEP	STR	VEL
SPO	995.9886 0.0000	8.7016 0.0690	8.3118 0.0400	0.1910 0.6621	0.4667 0.4945	2.2808 0.6843	1.8284 0.7673
IZO	8.7016 0.0690	4000.0000 0.0000	23.5731 0.0232	36.7544 0.0000	7.3770 0.1173	150.7656 0.0000	74.4857 0.0000
STN	8.3118 0.0400	23.5731 0.0232	3000.0000 0.0000	27.3947 0.0000	37.4065 0.0000	145.3793 0.0000	142.5799 0.0000
DEL	0.1910 0.6621	36.7544 0.0000	27.3947 0.0000	990.6200 0.0000	102.0077 0.0000	29.0470 0.0000	8.0524 0.0897
NEP	0.4667 0.4945	7.3770 0.1173	37.4065 0.0000	102.0077 0.0000	986.7597 0.0000	10.3055 0.0356	4.2206 0.3770
STR	2.2808 0.6843	150.7656 0.0000	145.3793 0.0000	29.0470 0.0000	10.3055 0.0356	4000.0000 0.0000	27.3823 0.0374
VEL	1.8284 0.7673	74.4857 0.0000	142.5799 0.0000	8.0524 0.0897	4.2206 0.3770	27.3823 0.0374	4000.0000 0.0000

Za vsak par nominalnih spremenljivk želimo preizkusiti domnevo, da med njima ne obstaja povezanost ($H_0 : f_{ij} = f'_{ij}$). Preizkusimo npr. domnevo o nepovezanosti med spolom in lastništvom delnic ali obveznic. Iz tabele 3.3 razberemo, da je vrednost statistike χ^2 , ki meri povezanost med spolom in lastništvom delnic ali obveznic, enaka 0,1910, ustrezna vrednost P pa 0,6621. Na podlagi izvedenega preizkusa ne moremo trditi, da obstaja povezanost med spolom in lastništvom delnic ali obveznic.

⁸ Pri izračunu statistike χ^2 za kontingenčne tabele 2x2 smo upoštevali Yatesov popravek.

Izhodišče multiple korespondenčne analize je Burtova tabela **B**, prikazana v tabeli 3.1. Na podlagi Burtove tabele formiramo matriko $\mathbf{D}_B^{-1/2}(\mathbf{P}_B - \mathbf{p}_B \mathbf{p}'_B) \mathbf{D}_B^{-1/2}$ (3.13). S spektralno dekompozicijo te matrike dobimo diagonalno matriko glavnih inercij \mathbf{D}_λ (3.19) in matriko standardiziranih koordinat glavnih komponent $\mathbf{\Gamma}$ (3.20, glej tabelo 3.4).

Tabela 3.4 Matrika standardiziranih koordinat glavnih komponent $\mathbf{\Gamma}$ - prikaz prvih 8 komponent (Vir: Tabela 3.1).

		1. GK	2. GK	3. GK	4. GK	5. GK	6. GK	7. GK	8. GK
1	SPOM	0.2833	-0.3499	-0.3346	-0.2082	0.9260	0.8113	0.2380	1.2187
2	SPOZ	-0.2503	0.3091	0.2956	0.1839	-0.8178	-0.7166	-0.2102	-1.0764
3	IZO1	0.5297	-0.7304	-2.3698	0.9389	1.8266	0.2578	-0.3234	-2.1682
4	IZO2	0.8673	-1.2831	0.2717	-1.2127	-0.7274	0.3296	-0.3045	0.4058
5	IZO3	-0.7951	-0.0023	0.6979	1.4122	-0.7277	-3.2824	-0.5225	1.2460
6	IZO4	-1.0921	0.9759	1.8700	2.0598	-0.1067	1.5464	0.5215	0.2357
7	IZO5	-0.5189	2.6674	-0.3220	-2.3828	0.0940	0.7243	1.0631	0.3570
8	STN1	-0.2211	-1.5150	3.6819	-0.8791	2.0921	1.5394	-0.3222	-0.2802
9	STN2	2.3124	0.1441	-0.1457	0.3341	-0.4626	-0.2785	0.2824	0.5715
10	STN3	-1.1598	0.2405	-0.5799	-0.3999	-0.2733	-0.6155	-0.5146	-0.4140
11	STN4	-0.6713	0.2144	-1.5434	3.5729	0.4628	3.6164	3.7146	1.3761
12	DEL1	1.9612	3.7897	1.1176	-0.0511	0.3476	-0.1110	-1.0630	0.6198
13	DEL2	-0.2700	-0.5217	-0.1538	0.0070	-0.0478	0.0153	0.1463	-0.0853
14	NEP1	2.7528	3.5792	2.1868	1.3589	-0.3632	0.2410	-1.5673	0.2781
15	NEP2	-0.2459	-0.3197	-0.1953	-0.1214	0.0324	-0.0215	0.1400	-0.0248
16	STR1	-1.4415	0.0318	0.4872	8.0912	2.0612	-2.2946	1.1803	0.4106
17	STR2	-1.8457	0.5639	-0.3656	-0.1372	-0.8502	0.9448	-1.0524	1.4191
18	STR3	-0.1470	-0.4361	1.4567	-0.9376	0.5912	-0.4826	1.4492	-0.0673
19	STR4	1.2993	-0.7329	-0.3378	0.3828	-2.5068	0.2011	-1.2929	-1.7316
20	STR5	1.9030	0.9023	-2.2738	-0.1653	2.2978	-0.0449	-0.3559	-0.1032
21	VEL1	2.5583	-2.2274	-0.5065	1.0672	-3.7634	0.5426	2.4530	2.6613
22	VEL2	0.9762	-0.7237	2.6423	0.8295	0.6800	1.8348	0.0091	-4.4328
23	VEL3	0.5311	-0.6222	0.5385	-0.4935	2.0259	-3.0927	0.2306	0.9911
24	VEL4	-0.5191	-0.3096	-0.1546	0.2776	0.3416	1.0227	-2.5862	0.8778
25	VEL5	-0.6731	1.4067	-0.7092	-0.5084	-0.6555	0.0003	1.8592	-0.9125

Število netrivialnih inercij kot tudi število glavnih komponent je $J - Q = 18$. Vrednosti glavnih inercij λ_j , odstotnih deležev glavnih inercij in kumulativnih odstotnih deležev glavnih inercij so prikazane v tabeli 3.5.

Deleži nekaj prvih glavnih inercij so praviloma relativno nizki. Kot pa smo že omenili v razdelku 3.4, Benzécri (1979) meni, da so analitično pomembne le tiste glavne komponente, katerih inercije λ_j so večje od $\frac{1}{Q}$, torej v našem primeru večje od $\frac{1}{7} = 0.1429$. Izračunajmo modificirane inercije, upoštevaje Benzécrijev izraz (3.21), in prikažimo obe vrsti inercij λ_j in $\tilde{\lambda}_j$, odstotne deleže modificiranih inercij $\tilde{\lambda}_j\%$ in kumulativne odstotne deleže modificiranih inercij $kum\tilde{\lambda}_j\%$ v tabeli 3.6.

Tabela 3.5 Prikaz glavnih inercij λ_j , odstotnih deležev glavnih inercij $\lambda_j\%$ (prikazani so tudi v histogramu) in kumulativnih odstotnih deležev glavnih inercij $kum\lambda_j\%$ (Vir: Tabela 3.1)

	λ_j	$\lambda_j\%$	$kum\lambda_j\%$	2	4	6	8	10
1	0.2500	9.72	9.72	*****				
2	0.2184	8.49	18.21	*****				
3	0.1847	7.18	25.40	*****				
4	0.1746	6.79	32.19	*****				
5	0.1610	6.26	38.45	*****				
6	0.1580	6.15	44.59	*****				
7	0.1542	6.00	50.59	*****				
8	0.1457	5.67	56.26	*****				
.....								
9	0.1409	5.48	61.74	*****				
10	0.1344	5.23	66.96	*****				
11	0.1290	5.02	71.98	*****				
12	0.1253	4.87	76.85	*****				
13	0.1223	4.76	81.61	*****				
14	0.1177	4.58	86.18	*****				
15	0.1060	4.12	90.31	*****				
16	0.0957	3.72	94.03	*****				
17	0.0869	3.38	97.41	*****				
18	0.0666	2.59	100.00	*****				
	2.5714							

Tabela 3.6 Prikaz inercij λ_j in $\tilde{\lambda}_j$, odstotnih deležev modificiranih inercij $\tilde{\lambda}_j\%$ (prikazani so tudi v histogramu) in kumulativnih odstotnih deležev modificiranih inercij $kum\tilde{\lambda}_j\%$ (Vir: Tabela 3.5).

	λ_j	$\tilde{\lambda}_j$	$\tilde{\lambda}_j\%$	$kum\tilde{\lambda}_j\%$	10	20	30	40	50
1	0.2500	0.015631	55.65	55.65	*****				
2	0.2187	0.007760	27.63	83.27	*****				
3	0.1847	0.002378	8.47	91.74	****				
4	0.1746	0.001372	4.88	96.62	**				
5	0.1610	0.000450	1.60	98.22	*				
6	0.1580	0.000313	1.12	99.34	*				
7	0.1542	0.000174	0.62	99.96					
8	0.1457	0.000011	0.04	100.00					
	0.028089								

Kot lahko vidimo iz tabele 3.6, se je število glavnih inercij sedaj skrčilo od 18 na vsega 8, pri čemer pa že s prvima dvema glavnima komponentama predstavimo kar 83,27% skupne inercije.

Izračunali smo že matriko standardiziranih vrednosti glavnih komponent Γ (tabela 3.4). Sedaj moramo transformirati prvih osem stolpcev matrike Γ v ustrezne stolpce matrike vrednosti glavnih komponent Y (3.22) (v "nestandardizirane" glavne komponente). Matrika vrednosti glavnih komponent Y je prikazana v tabeli 3.7.

Tabela 3.7 Matrika vrednosti glavnih komponent Y (Vir: Tabela 3.4 in 3.6).

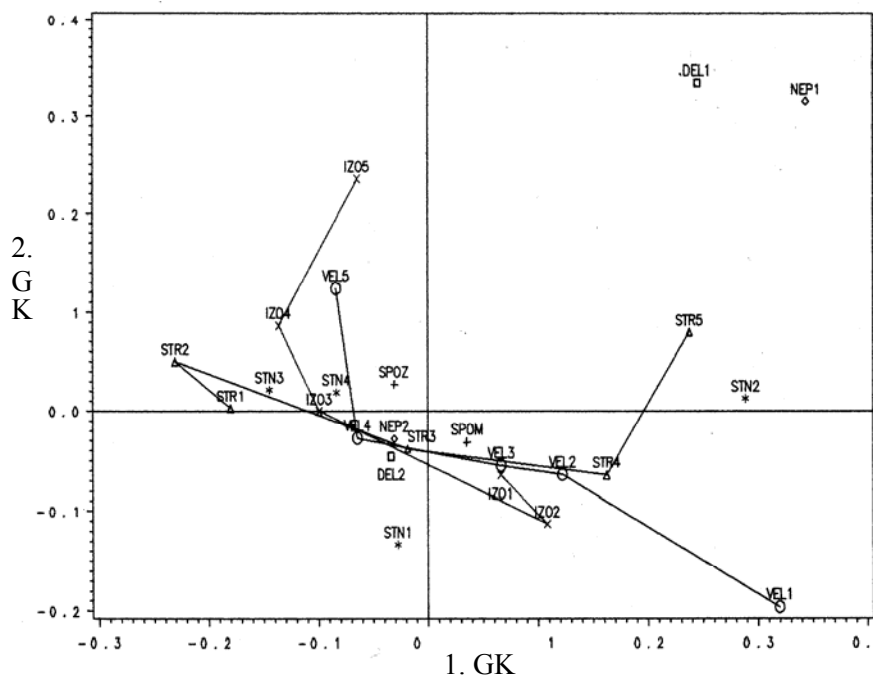
	1. GK	2. GK	3. GK	4. GK	5. GK	6. GK	7. GK	8. GK
1 SPOM	0.0354	-0.0308	-0.0163	-0.0077	0.0196	0.0144	0.0031	0.0041
2 SPOŽ	-0.0313	0.0272	0.0144	0.0068	-0.0173	-0.0127	-0.0028	-0.0036
3 IZO1	0.0662	-0.0643	-0.1156	0.0348	0.0387	0.0046	-0.0043	-0.0072
4 IZO2	0.1084	-0.1130	0.0132	-0.0449	-0.0154	0.0058	-0.0040	0.0014
5 IZO3	-0.0994	-0.0002	0.0340	0.0523	-0.0154	-0.0581	-0.0069	0.0041
6 IZO4	-0.1365	0.0860	0.0912	0.0763	-0.0023	0.0274	0.0069	0.0008
7 IZO5	-0.0649	0.2350	-0.0157	-0.0883	0.0020	0.0128	0.0140	0.0012
8 STN1	-0.0276	-0.1335	0.1796	-0.0326	0.0444	0.0272	-0.0043	-0.0009
9 STN2	0.2891	0.0127	-0.0071	0.0124	-0.0098	-0.0049	0.0037	0.0019
10 STN3	-0.1450	0.0212	-0.0283	-0.0148	-0.0058	-0.0109	-0.0068	-0.0014
11 STN4	-0.0839	0.0189	-0.0753	0.1323	0.0098	0.0640	0.0491	0.0046
12 DEL1	0.2452	0.3338	0.0545	-0.0019	0.0074	-0.0020	-0.0140	0.0021
13 DEL2	-0.0338	-0.0460	-0.0075	0.0003	-0.0010	0.0003	0.0019	-0.0003
14 NEP1	0.3442	0.3153	0.1066	0.0503	-0.0077	0.0043	-0.0207	0.0009
15 NEP2	-0.0307	-0.0282	-0.0095	-0.0045	0.0007	-0.0004	0.0018	-0.0001
16 STR1	-0.1802	0.0028	0.0238	0.2997	0.0437	-0.0406	0.0156	0.0014
17 STR2	-0.2308	0.0497	-0.0178	-0.0051	-0.0180	0.0167	-0.0139	0.0047
18 STR3	-0.0184	-0.0384	0.0710	-0.0347	0.0125	-0.0085	0.0191	-0.0002
19 STR4	0.1624	-0.0646	-0.0165	0.0142	-0.0532	0.0036	-0.0171	-0.0058
20 STR5	0.2379	0.0795	-0.1109	-0.0061	0.0487	-0.0008	-0.0047	-0.0003
21 VEL1	0.3198	-0.1962	-0.0247	0.0395	-0.0798	0.0096	0.0324	0.0089
22 VEL2	0.1220	-0.0638	0.1289	0.0307	0.0144	0.0325	0.0001	-0.0148
23 VEL3	0.0664	-0.0548	0.0263	-0.0183	0.0430	-0.0547	0.0030	0.0033
24 VEL4	-0.0649	-0.0273	-0.0075	0.0103	0.0072	0.0181	-0.0342	0.0029
25 VEL5	-0.0841	0.1239	-0.0346	-0.0188	-0.0139	0.0000	0.0246	-0.0030

Vrednosti glavnih komponent bodo rabile kot koordinate projekcij profilov-točk na optimalni podprostor izbrane razsežnosti.

3.5 Prikaz profilov v dvorazsežnem razsevnom grafikonu

Profili so v geometrijskem smislu idealno predstavljeni kot točke v večrazsežnem prostoru. Število razsežnosti takšnega prostora je enako skupnemu številu glavnih komponent. V praksi pa se pogosto odločimo, da prikažemo njihov položaj v dvorazsežnem razsevnom grafikonu. Če je skupno število glavnih komponent večje od dve pomeni takšen prikaz položaja

profilov-točk seveda le približek njihovega realnega položaja, položaj točk v prvotnem prostoru je aproksimiran z njihovimi ortogonalnimi projekcijami na optimalni dvorazsežni podprostor. V tem primeru so v korespondenčni analizi koordinate točk vrednosti prvih dveh glavnih komponent. V našem primeru so te vrednosti predstavljene v prvih dveh stolpcih matrike vrednosti glavnih komponent Y (tabela 3.7), položaj profilov-točk pa prikazuje slika 3.1.



Slika 3.1 Prikaz položaja profilov-točk v optimalni ravnini (Vir: tabela 3.7)

Pojasnimo najprej položaj profilov-točk za vsako spremenljivko posebej (t.i. “within-set distances”), pri čemer vzemimo kot osnovo za oceno položaja profilov koordinatno izhodišče, ki predstavlja položaj povprečnega profila-centroida.

Za točki, ki predstavljata položaj profilov spremenljivke spol, je značilno, da ležita relativno blizu koordinatnega izhodišča, kar kaže na to, da se anketirani moški (SPOM) oz. ženske (SPOŽ) s stališča ostalih spremenljivk med seboj

bistveno ne razlikujejo. To je razvidno tudi iz matrike vrstičnih profilov, kajti profili SPOM, SPOŽ in povprečni profil se le malo razlikujejo med seboj. Poglejmo še vrednosti statistike χ^2 in pripadajoče vrednosti P (tabela 3.3) za vse možne pare, ki jih tvori spremenljivka spol z ostalimi spremenljivkami. Povezanost odkrijemo le med spolom in stanovanjskimi razmerami.

Za vsako dihotomno spremenljivko (npr. spol) je značilno, da njena profila točki določata daljico, ki poteka skozi koordinatno izhodišče. Profil-točka skupine z večjim deležem enot leži bliže koordinatnemu izhodišču in obratno.

Na podlagi položaja točk, ki predstavljajo profile spremenljivke najvišja stopnja dosežene šolske izobrazbe, ugotavljamo: a) da sta profila obeh najnižjih stopenj izobrazbe - brez dokončane osnovne šole in dokončana osnovna šola (IZO1 in IZO2) precej podobna; b) da imata profila obeh najvišjih stopenj izobrazbe - dokončana srednja šola in vsaj nedokončana fakulteta (IZO4 in IZO5) nasproten položaj glede na najnižji stopnji izobrazbe (z vidika koordinatnega izhodišča); c) da izrazito izstopa položaj profila vsaj nedokončana fakulteta v smeri osi 2. glavne komponente. Med posameznimi profili izobrazbe torej obstajajo precejšnje razlike z vidika ostalih proučevanih spremenljivk. To trditev podkrepljuje tudi neposredna primerjava vrednosti profilov izobrazbe, kjer opazimo precejšnje razlike (gledano z vidika vseh šestih spremenljivk hkrati, pri čemer pa so pri posamezni spremenljivki razlike lahko tudi relativno majhne). Tudi vrednosti statistike χ^2 odkrivajo povezanost spremenljivke najvišja stopnja dosežene šolske izobrazbe s štirimi izmed ostalih spremenljivk

Na podlagi položaja točk, ki predstavljajo profile spremenljivke stanovanjske razmere, ugotavljamo podoben položaj najemnikov (STN3) in oseb, ki imajo brezplačno bivanje (STN4), in povsem nasproten položaj lastnikov (STN2). Položaj omenjenih profilov-točk je predvsem izraz vrednosti 1. glavne komponente. Položaj hipotečnih dolžnikov (STN1) pa je predvsem izražen z vrednostjo 2. glavne komponente. Dodatna potrditev teh sklepov z vidika neposredne primerjave vrednosti profilov in vrednosti statistike χ^2 je analogna kot pri že obravnavanih spremenljivkah, zato je pri tej in pri preostalih spremenljivkah ne podajamo.

Na podlagi položaja točk, ki predstavljajo profile spremenljivke lastništvo

delnic in obveznic oz. profile spremenljivke lastništvo nepremičnin, ugotavljamo, da so v obeh primerih profili oseb brez tovrstnega imetja (DEL2 in NEP2) blizu koordinatnega izhodišča, profili oseb, ki imajo tovrstno imetje (DEL1 in NEP1), pa močno oddaljeni od njega. V našem primeru ima delnice ali obveznice le 8.2% oseb, nepremičnine (brez stanovanja, v katerem biva) pa le 12.1% oseb (deleži so izračunani na podlagi tabele 3.1). Zato je razumljivo, da je vpliv oseb brez omenjene imovine na položaj povprečnega profila-centroida prevladujoč, centroid je blizu njihovih profilov-točk. Očitno je torej, da med lastniki delnic in obveznic in ostalimi osebami obstajajo velike razlike z vidika ostalih šestih spremenljivk. Podobno tudi med lastniki nepremičnin in ostalimi osebami obstajajo velike razlike z vidika ostalih šestih spremenljivk.

Položaj profilov-točk spremenljivke starost (STR1 do STR5) in profilov-točk spremenljivke velikost kraja bivanja (VEL1 do VEL5) je opredeljen predvsem z vrednostmi 1. glavne komponente. Med posameznimi starostnimi profili obstajajo precejšnje razlike glede na ostalih šest spremenljivk. Podobno tudi med posameznimi profili spremenljivke velikost kraja bivanja obstajajo precejšnje razlike z vidika ostalih šestih spremenljivk.

Med proučevanimi spremenljivkami imajo spremenljivke najvišja stopnja dosežene šolske izobrazbe, starost in velikost kraja bivanja višjo stopnjo merske lestvice od nominalne. V takšnih primerih lahko povežemo profile-točke za vsako spremenljivko posebej z daljicami, primerjava poteka dobljenih lomljenih črt pa naj bo podlaga pri boljši vsebinski razlagi povezav med spremenljivkami. V našem primeru na podlagi slike 3.1 ugotavljamo, da obstaja določena podobnost v poteku vseh treh lomljenih črt, kar kaže na vsebinsko povezanost omenjenih treh spremenljivk. Slednje potrjujejo tudi vrednosti statistike χ^2 , ki so za vse možne pare omenjenih spremenljivk statistično značilne.

V primeru multiple korespondenčne analize je dopustno primerjanje medsebojnega položaja profilov-točk, ki se nanašajo na različne spremenljivke (t.i. "between-set distances"), čeprav medsebojna razdalja tudi v tem primeru ni definirana. Tako lahko primerjamo položaj dveh profilov, ki pripadata dvema skupinama različnih spremenljivk, in imata podoben položaj v optimalnem podprostoru. Takšni dve skupini sta podobni, ker se približno ujemajo njune

vrednosti profilov glede na preostale spremenljivke.

Na podlagi slike 3.1 lahko oblikujemo npr. naslednje štiri skupine profilov:

- a) profil oseb z dokončano srednjo šolo (IZO4), profil najemnikov stanovanja (STN3), profil oseb starih od 19 do 24 let (STR1) in profil oseb starih od 25 do 34 let (STR2);
- b) profil lastnikov stanovanja (STN2) in profil oseb starejših od 65 let (STR5);
- c) profil lastnikov delnic ali obveznic (DEL1) in profil lastnikov nepremičnin (NEP1);
- d) profila oseb obeh spolov (SPOM in SPOŽ), profil oseb, ki niso lastniki delnic ali obveznic (DEL2), profil oseb, ki niso lastniki nepremičnin (NEP2), profil oseb starih med 35 in 49 let (STR3) in profil oseb, ki živijo v mestih, ki imajo od 100000 in 500000 prebivalcev.

Na podlagi tako oblikovanih skupin profilov ugotavljamo naslednje:

- da je za del oseb značilno, da se ujemajo v pretežnem delu naslednjih lastnosti: stare so od 19 do 34 let, bivajo v najemniških stanovanjih in imajo dokončano srednjo šolo;
- da so osebe, ki so lastniki stanovanja, pogosto stare nad 65 let;
- da so osebe, ki so lastniki delnic ali obveznic pogosto tudi lastniki nepremičnin;
- da je za del oseb značilno, da se ujemajo v pretežnem delu naslednjih lastnosti: so obeh spolov, stare so od 35 do 49 let, bivajo v mestih, ki imajo od 100000 do 500000 prebivalcev in niso lastniki niti delnic niti obveznic niti nepremičnin (razen stanovanja).

Zavedati pa se moramo, da je podano tolmačenje lahko le toliko "dobro", kolikor je dobra aproksimacija prvotnega prostora v optimalnem dvorazsežnem podprostoru. Ker je v našem primeru kumulativni delež inercije prvih dveh glavnih komponent v skupni inerciji 83,27% (glej tabelo 3.6), je sklepanje na podlagi dvorazsežnega razsevnega grafikona nujno pomanjkljivo. ■

4 Literatura

- Andrews, D. F. 1972. "Plots of high dimensional data." *Biometrics* 28:125-136.
- Benzécri, J.-P. 1979. "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum a [BIN.MULT]." *Cahiers de L'analyse des Données* 4:377-378.
- Benzécri, J.-P. et al. 1980. *L'Analyse des Données, L'Analyse des Correspondances*. Paris: Dunod.
- Burt C.: The factorial analysis of qualitative data, *British Journal of Psychology*, vol3, str. 166-185, 1950
- Fisher R.A.: The precision of discriminant functions, *Annals of Eugenics*, vol. 10, str. 422-429, 1940
- Hirschfeld H.O.: A connection between correlation and contingency, *Cambridge Philosophical Society Proceedings*, vol. 31, str. 520-524, 1935
- Hill M.O.: Correspondence Analysis: A neglected multivariate method, *Appl. Statistics*, vol. 23, str. 340-354, 1974
- Greenacre, M. J. 1981. "Practical correspondence analysis." Pp.119-146, in *Interpreting Multivariate Data*, edited by V. Barnett. Chichester, U.K.: Wiley.
- Greenacre M.J.: *Theory and Application of Correspondence Analysis*, Academic Press, 1984
- Greenacre, M. J., and T. Hastie. 1987. "The geometric interpretation of correspondence analysis." *Journal of the American Statistical Association* 82:437-447.
- Greenacre, M.J.: *Correspondence analysis in practice*, Academic Press, 1993.
- Greenacre, M.J. (1994). Multiple and joint correspondence analysis, in M.J. Greenacre and J. Blasius (eds.), *Correspondence Analysis in the Social Sciences*, pp. 141-161. London: Academic Press.
- Guttman L.: The quantification of a class of attributes: A theory and method of scale construction v Horst P. ured.: *The Prediction of Personal Adjustment*, str. 319-348. Social Science Research Council, New York, 1941
- Kendall D.G.: Seriation from abundance matrices, str. 215-252 v Hodson C.R., Kendall D.G. in Tautu P.(ed.): *Mathematics in the Archaeological and Historical Sciences*, Edinburg University Press, Edinburg, 1971
- Košmelj B. et al.: *Statistični terminološki slovar*, Statistično društvo Slovenije, Slovenska akademija znanosti in umetnosti, Ljubljana, 2001
- Lebart, L., and Y. Houzel van Effenterre. 1980. "Le Systeme d'Enquetes sur les Aspirations des Français: Une Breve Présentation." *Consommation* 1:3-25.

- Lebart L., A. Morineau, and K. M. Warwick. 1984. *Multivariate Descriptive Statistical Analysis, Correspondance Analysis and Related Techniques for Large Matrices*. New York: Wiley.
- Pearson K. 1901. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine* 2:559-572.
- Rovan, J. Andrewsove krivulje v korespondenčni analizi. V: Ferligoj, A. (ur.), Jug, J (ur.). *Blejsko metodološko srečanje '91 : zbornik referatov IX. Strokovnega sestanka Sekcije za metodologijo in statistiko Jugoslovanskega združenja za sociologijo*, (Zbirka Metodološki zvezki, št. 8). Ljubljana: Fakulteta za družbene vede, Inštitut za družbene vede, 1991, str. 64-97.
- Rovan, J. The role of Andrews' curves correspondence analysis. V: *SEGUI'91 : proceedings of the SAS European users group : international conference June 11-14, 1991*. Amsterdam: SAS, 1991, str. 460-474.
- Rovan, J. *Andrewsove krivulje v korespondenčni analizi : doktorska disertacija*. Ljubljana: [s.n.], 1991. III, 253 str., ilustr.
- Rovan, J. Visualizing solutions in more than two dimensions. V: Greenacre, M. (ur.), Blasius, J. (ur.). *Correspondence analysis in the social sciences : recent developments and applications*. San Diego...[etc.]: Academic Press, cop. 1994, str. 210-229.
- Rovan, J., Mramor, D., Horvat, I. *Multiple correspondence analysis in bank services : predstavljeno na posvetovanju "SAS European users group international" SEUGI, Praha 9-12 June 1998*. Praha, 1998: SEUGI. [11 str.].
- Horvat I., Rován J., Ferligoj A.: Segmentation of Slovenian retail bank market. Zbornik referatov, SAS European Users Group International, SEUGI'99, Den Haag, 1999. CD-rom.
- Rovan J., Urbančič-Rovan V., Slak M.: Screening procedures in human medicine – evaluation of results by multiple correspondence analysis. Zbornik referatov, SAS European Users Group International, SEUGI'2000, Dublin, 2000. CD-rom.
- Tenenhaus M. in Young F.W.: An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data, *Psychometrica*, vol. 50, str. 91-119, 1985
- du Toit S. H. C., A. G. W. Steyn, and R. H. Stumpf. 1986. *Graphical Exploratory Data Analysis*. New York: Springer.
- Wang P. C. C., ed. 1978. *Graphical Representation of Multivariate Data*. New York: Academic Press.